

Optimal Design of the Shiryaev–Roberts Chart: Give Your Shiryaev–Roberts a Headstart

Aleksey S. Polunchenko

Abstract We offer a numerical study of the effect of headstarting on the performance of a Shiryaev–Roberts (SR) chart set up to control the mean of a normal process. The study is a natural extension of that previously carried out by Lucas & Crosier (1982) for the CUSUM scheme. The Fast Initial Response (FIR) feature exhibited by a headstarted CUSUM turns out to be also characteristic of an SR chart (re-)started off a positive initial score. However, our main result is the observation that a FIR SR with a carefully designed *optimal* headstart is not just faster to react to an initial out-of-control situation, it is nearly *the* fastest *uniformly*, i.e., assuming the process under surveillance is equally likely to go out of control effective any sample number. The performance improvement is the greater, the fainter the change. We explain the optimization strategy, and tabulate the optimal initial score, control limit, and the corresponding “worst possible” out-of-control Average Run Length (ARL), considering mean-shifts of diverse magnitudes and a wide range of levels of the in-control ARL.

1 Introduction

The general theme of this work is the optimal design of the Shiryaev–Roberts (SR) chart originally proposed by Shiryaev (1961, 1963) and Roberts (1966), and later generalized by Moustakides et al. (2011). Set up to detect a possible change in the baseline mean of a series of independent samples X_1, X_2, \dots drawn from a normal unit-variance population at regular time intervals, the classical SR chart involves sequential evaluation of the SR statistic $\{R_n\}_{n \geq 0}$ using the recurrence $R_n = (1 + R_{n-1}) \exp\{S_n\}$, $n = 1, 2, \dots$, with $R_0 = 0$, and where the quantity

A.S. Polunchenko

Department of Mathematical Sciences, State University of New York at Binghamton, Binghamton, New York 13902–6000, USA e-mail: aleksey@binghamton.edu

$$S_n \triangleq \mu \left(X_n - \frac{\mu}{2} \right) \quad (1)$$

is a numerical score that captures the severity of the deviation of the n -th sample point X_n from the target mean-value in either direction; the score function S_n assumes that the intended (target) mean-value of the data is zero, but it is anticipated to change abruptly and permanently to a known off-target value $\mu \neq 0$. The n -th observation X_n might represent a single reading or the average of a batch of observations from a designated routine sampling plan. The chart triggers an alarm at the first stage, \mathcal{S}_A , such that $R_{\mathcal{S}_A} \geq A$, where $A > 0$ is a control limit (detection threshold) set in advance in accordance with the desired level of the false alarm risk; more formally, $\mathcal{S}_A \triangleq \min\{n \geq 1: R_n \geq A\}$, where $A > 0$ is given. Hence the process $\{X_n\}_{n \geq 1}$ is considered to be in control until stage \mathcal{S}_A . The random variable, \mathcal{S}_A , referred to as the run length, is the stage at which sampling stops and appropriate action is taken. A brief account of the history of the SR chart was recently offered by Pollak (2009). For an up-to-date summary of the classical as well as generalized SR charts' optimality properties, see, e.g., Polunchenko & Tartakovsky (2012).

Though nowhere nearly as known and as widespread as Page's (1954) celebrated CUSUM "inspection scheme", the SR chart did receive some attention in the applied literature. One of the earliest investigations of the chart's characteristics is due to Roberts (1966), who offered a performance comparison of the chart against a host of other statistical process control procedures, including the CUSUM scheme and the EWMA chart (also introduced by Roberts, 1959). A similar type of SR-vs-CUSUM comparison (but with respect to a different criterion and for a different data model) was also later performed by Mevorach & Pollak (1991). See also, e.g., Tartakovsky & Ivanova (1992), Tartakovsky et al. (2009), and Moustakides et al. (2009). Certain data-analytic advantages of the chart over the CUSUM scheme were pointed out by Kenett & Pollak (1996). Kenett & Pollak (1986) provided an example of an application of the SR chart in the area of software reliability.

In the (more theoretical) area of quickest change-point detection, the SR chart received far more attention. To a large extent this is due to the fundamental work of Shiryaev (1961, 1963) who proved that the chart solves a particular Bayesian version of the quickest change-point detection problem; see also Girshick & Rubin (1952). The chart then remained unnoticed until recently Pollak & Tartakovsky (2009) and Shiryaev & Zryumov (2009) discovered that it solves yet another so-called *multi-cyclic* or *generalized* Bayesian version of the quickest change-point detection problem; the multi-cyclic setup is instrumental in such applications as cybersecurity (see, e.g., Tartakovsky et al., 2013), financial monitoring (see, e.g., Pepelyshev & Polunchenko, 2016), and economic design of control charts. This brought the SR chart back into the spotlight. Polunchenko et al. (2016) performed a robustness analysis of the SR chart's multi-cyclic capabilities when the post-change distribution involves a misspecified parameter. Moustakides et al. (2011) observed that by starting the SR statistic $\{R_n\}_{n \geq 0}$ off a positive initial value, i.e., setting $R_0 = r > 0$, the SR chart can be made nearly the best (in the minimax sense of Pollak, 1985). Roughly, this means the SR chart is almost the fastest to react to a change in the observations' distribution when the corresponding unknown change-point is equally likely to be any

point in time; see Section 2 for a formal definition. As a matter of fact Polunchenko & Tartakovsky (2010) and Tartakovsky & Polunchenko (2010) demonstrated that in two specific change-point scenarios the SR chart with a carefully designed headstart is *the* fastest (in the sense of Pollak, 1985). This result was then extended by Tartakovsky et al. (2012) who proved that the SR chart whose headstart is selected in a specific fashion is almost the best one can do (again, in the sense of Pollak, 1985) asymptotically, as the false alarm risk tends to zero, in a general change-point scenario.

In spite of the aforementioned strong theoretically established optimality properties of the SR chart, and the fact that no such properties are exhibited by either the CUSUM scheme or the EWMA chart, applications of the SR chart in quality control remain essentially nonexistent. In part, this may be due to the lack of existing resources with pre-computed, for a variety of cases, optimal headstart and control limit values. To the best of our knowledge, the work of Tartakovsky et al. (2009) and that of Polunchenko & Sokolov (2014) have heretofore been the only sources with such data (computed assuming the observations are exponential). This work’s goal is to optimize the SR chart for yet another model, namely, the standard Gaussian model widely used in the quality control literature as a testbed for charts’ performance analysis. The specific optimization strategy is presented in Section 2. The optimization itself is carried out in Section 3 using the numerical framework developed by Moustakides et al. (2011) and then improved upon by Polunchenko et al. (2014b, 2014a). The obtained optimal headstart and control limit values are reported in Section 3 as well. Conclusions follow in Section 4.

2 The Shiryaev–Roberts Chart, Its Properties and Optimization

To control the mean of a standard Gaussian process, the headstarted tweak of the classical SR chart proposed by Moustakides et al. (2011) operates by sequentially updating the statistic $\{R_n^r\}_{n \geq 0}$ via the recurrence

$$R_n^r = (1 + R_{n-1}^r) \exp\{S_n\}, \quad n = 1, 2, \dots \text{ with } R_0^r = r \geq 0, \quad (2)$$

where S_n is the score function defined in (1); the initial score $R_0^r = r \geq 0$ is a design parameter also referred to as the headstart, which is the original terminology of Lucas & Crosier (1982) who suggested to headstart the CUSUM scheme. The corresponding run length is as follows:

$$\mathcal{S}_A^r \triangleq \min\{n \geq 1 : R_n^r \geq A\}, \quad (3)$$

where $A > 0$ is the control limit (detection threshold) selected in advance so as to keep the chart’s false alarm characteristics tolerably low. Note that if $r = 0$ then the chart is the classical SR chart (with no headstart) of Shiryaev (1961, 1963) and Roberts (1966). For this reason Tartakovsky et al. (2012) coined the term “*Generalized SR chart*” (or the GSR chart for short) to refer to the headstarted SR chart

defined by (2) and (3). It is also worth reiterating that the score function (1)—and hence also the statistic (2)—are indifferent to the direction of the mean shift, i.e., the sign of $\mu \neq 0$ is irrelevant.

It has been the custom in the quality control literature to assess the operating characteristics of a control chart, with run length T , by means of only two indices: the in-control Average Run Length (ARL) and the out-of-control ARL. In this work, we shall adapt the (more exhaustive) approach used in the quickest change-point detection literature. Let \mathbb{P}_k (\mathbb{E}_k) denote the probability measure (expectation) induced by the data $\{X_n\}_{n \geq 1}$ assuming the change-point is at time moment $k = 0, 1, 2, \dots, \infty$, i.e., assuming the process $\{X_n\}_{n \geq 1}$ is in-control until sample number k inclusive, and is out-of-control starting from sample number $k + 1$ onward. The notation $k = 0$ ($k = \infty$) is to be understood as the case when the process under surveillance is out of control *ab initio* (never, respectively).

The in-control characteristics of a control chart T are usually gauged by virtue of the Average Run Length (ARL) to false alarm $\text{ARL}(T) \triangleq \mathbb{E}_\infty[T]$ which is the average number of samples taken by the chart before an *erroneous* out-of-control signal is given; this is precisely what is known in the quality control literature as the *in-control* ARL. It is apparent that the higher the ARL to false alarm, the lower the level of the false alarm risk. For the GSR chart, the general inequality $\text{ARL}(\mathcal{S}_A^r) \geq A - r$ can be used to design $A > 0$ and $r \in [0, A]$ so as to have $\text{ARL}(\mathcal{S}_A^r)$ no lower than a desired margin $\gamma > 1$. It is of note that this inequality holds in general, whatever the statistical structure of the observations be. A more accurate result is the asymptotic (as $A \rightarrow +\infty$) approximation $\text{ARL}(\mathcal{S}_A^r) \approx A/\xi - r$, which is actually known to be quite accurate even if $A > 0$ is not high; see, e.g., (Pollak, 1987, Theorem 1) or Tarakovsky et al. (2012). Here ξ denotes the so-called “limiting average exponential overshoot”—a model-dependent constant (taking values between 0 and 1) computable using nonlinear renewal-theoretic methods; see, e.g., Woodroffe (1982). For the Gaussian model considered in this work it follows, e.g., from (Woodroffe, 1982, Example 3.1, pp. 32–33), that the following formula can be used:

$$\xi = \frac{2}{\mu^2} \exp \left\{ -2 \sum_{m=1}^{\infty} \frac{1}{m} \Phi \left(-\frac{\mu}{2} \sqrt{m} \right) \right\}, \quad (4)$$

where

$$\Phi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

is the standard Gaussian cumulative distribution function. Note from the foregoing formula that ξ is an even function of $\mu \neq 0$. The formula was put to use by Woodroffe (1982) who computed ξ for various values of $\mu > 0$; see (Woodroffe, 1982, Table 3.1, p. 33) for the obtained results.

To quantify the capabilities of a control chart T when the process is no longer in control, Pollak (1985) suggested to use the “worst-case” (Supremum) Average Detection Delay (SADD), conditional on no false alarm having been sounded. Formally,

$$\text{SADD}(T) \triangleq \max_{0 \leq k < \infty} \text{ADD}_k(T),$$

where $\text{ADD}_k(T) \triangleq \mathbb{E}_k[T - k | T > k]$, $k = 0, 1, 2, \dots$. Incidentally, the limiting ADD value $\lim_{k \rightarrow \infty} \text{ADD}_k(T)$ is known in the quality control literature as the *steady-state* ARL.

Pollak’s (1985) criterion has a simple interpretation: for any fixed but finite $k = 0, 1, 2, \dots$, the condition $T > k$ guarantees that it is an actual detection (not a false alarm), so that each $\text{ADD}_k(T)$ is the average number of samples it takes the chart past the change-point k to realize the process is not in control anymore, and because k is unknown, it is reasonable to assume it equally likely to be any number $(0, 1, 2, \dots)$ and consider the worst possible case, i.e., take the maximal of the $\text{ADD}_k(T)$ ’s. For the CUSUM scheme with no headstart and for the classical SR chart (also headstart-free) it can be shown that $k = 0$ is when the ADD is the highest, i.e., $\text{SADD}(T) = \text{ADD}_0(T)$. As a result, it suffices to restrict attention to just $\text{ADD}_0(T)$, and it is this quantity that the quality control community calls the out-of-control ARL. However, things are not as simple when the chart has a positive headstart, and it is no longer obvious which of the delays $\text{ADD}_k(\mathcal{S}_A^r)$ ’s for $k = 0, 1, 2, \dots$ is the highest. As a matter of fact we shall see in the next section that the “bump” of the sequence $\{\text{ADD}_k(\mathcal{S}_A^r)\}_{k \geq 0}$ has a highly unpredictable behavior in terms of its location on the time axis.

Let $\Delta(\gamma) \triangleq \{T : \text{ARL}(T) \geq \gamma\}$ be the class of control charts (identified with a generic run length T) whose ARL to false alarm is at least as high as a desired pre-set level $\gamma > 1$. Pollak’s (1985) minimax change-point detection problem consists in finding $T_{\text{opt}} \in \Delta(\gamma)$ such that $\text{SADD}(T_{\text{opt}}) = \min_{T \in \Delta(\gamma)} \text{SADD}(T)$ for any given $\gamma > 1$. In general, this problem is still an open one, although there has been a continuous effort to solve it. To that end, for at least two specific data models, the answer was shown to be the GSR chart with “finetuned” threshold and headstart values; see Polunchenko & Tartakovsky (2010) and Tartakovsky & Polunchenko (2010). Moreover, for a general data model, the GSR chart (properly optimized) was also shown (by Tartakovsky et al. 2012) to solve Pollak’s (1985) problem asymptotically as $\gamma \rightarrow +\infty$. Specifically, this means that if A and r are selected so that $\text{ARL}(\mathcal{S}_A^r) \geq \gamma$ with $\gamma > 1$ given, i.e., $\mathcal{S}_A^r \in \Delta(\gamma)$, then

$$\text{SADD}(\mathcal{S}_A^r) - \min_{T \in \Delta(\gamma)} \text{SADD}(T) \searrow 0 \text{ as } \gamma \rightarrow +\infty, \quad (5)$$

provided, however, that $r/A \rightarrow 0$ as $A \rightarrow +\infty$; see Tartakovsky et al. (2012), who also supply a high-order large- γ expansion of $\text{SADD}(\mathcal{S}_A^r)$. The foregoing is a strong optimality property known in the literature on change-point detection as asymptotic minimax optimality of order three, or asymptotic near minimaxity. It is noteworthy that the CUSUM chart, whether headstarted or not, does not have such strong “nearly-best” detection capabilities. Moreover, nor does the EWMA chart. Hence, our interest in the GSR chart. To provide an idea as to the difference made by a positive headstart, we remark that the classical SR chart (with zero headstart) is asymptotically (as $\gamma \rightarrow +\infty$) minimax of order two, i.e., the difference $\text{SADD}(\mathcal{S}_A) - \min_{T \in \Delta(\gamma)} \text{SADD}(T)$ goes to a positive constant as $\gamma \rightarrow +\infty$. More-

over, since the constant is the higher, the fainter the change, giving an SR chart a positive headstart is especially beneficial when the out-of-control behavior of the process differs from its in-control behavior only slightly.

Yet another strong optimality property of the GSR chart is its exact multi-cyclic or generalized Bayesian optimality. Specifically, Pollak & Tartakovsky (2009) and Shiryaev & Zryumov (2009) proved that the classical SR chart (with no headstart) minimizes the so-called Integral ADD

$$\text{IADD}(T) \triangleq \sum_{k=0}^{\infty} \mathbb{E}_k[\max\{0, T - k\}], \quad (6)$$

and the so-called Relative IADD (RIADD)

$$\text{RIADD}(T) \triangleq \text{IADD}(T) / \text{ARL}(T) = \sum_{k=0}^{\infty} \frac{\mathbb{P}_{\infty}(T > k)}{\text{ARL}(T)} \text{ADD}_k(T), \quad (7)$$

both inside the class $\Delta(\gamma)$ defined above, for any $\gamma > 1$. The meaning of this result can be explained by analyzing the structure of the definition (7) of $\text{RIADD}(T)$. Specifically, on the one hand, the latter can be viewed as being the k -average of the delays $\mathbb{E}_k[\max\{0, T - k\}]$, $k = 0, 1, 2, \dots$, assuming that change-point k has an improper uniform distribution on the set $\{0, 1, 2, \dots\}$. The improper uniformity of the change-point is a core assumption of the generalized Bayesian change-point detection problem. On the other hand, $\text{RIADD}(T)$ can also be regarded as the k -average of the $\text{ADD}_k(T)$'s assuming that the probability mass function of k is given by the ratio $\mathbb{P}_{\infty}(T > k) / \text{ARL}(T)$, $k = 0, 1, 2, \dots$; note that $\mathbb{P}_k(T > k) \equiv \mathbb{P}_{\infty}(T > k)$ for any $k = 0, 1, 2, \dots$, and that $\text{ARL}(T) = \sum_{k=0}^{\infty} \mathbb{P}_{\infty}(T > k)$. For yet another, viz. multi-cyclic interpretation, see Pollak & Tartakovsky (2009).

The RIADD-optimality of the classical SR chart was generalized in (Polunchenko & Tartakovsky, 2010, Lemma 1) where it was shown that the GSR chart, whose control limit $A > 0$ and headstart $r \geq 0$ are such that $\text{ARL}(\mathcal{S}_A^r) \geq \gamma$ for a given $\gamma > 1$, minimizes the so-called Stationary ADD (STADD)

$$\text{STADD}(T) \triangleq (r \text{ADD}_0(T) + \text{IADD}(T)) / (\text{ARL}(T) + r) \quad (8)$$

inside class $\Delta(\gamma)$, for any $\gamma > 1$; recall that $\text{IADD}(T)$ is as in (6). Formally, for any $\gamma > 1$, and any $A > 0$ and $r \geq 0$, it holds true that $\text{STADD}(\mathcal{S}_A^r) = \min_{T \in \Delta(\gamma)} \text{STADD}(T)$, provided that $\text{ARL}(\mathcal{S}_A^r) \geq \gamma$ is satisfied. Also, observe that $\text{STADD}(\mathcal{S}_A^r)$ reduces to $\text{RIADD}(\mathcal{S}_A^r)$ when $r = 0$. It is also of note that $\text{STADD}(T)$ is not the same as the limit $\lim_{k \rightarrow \infty} \text{ADD}_k(T)$.

An important “by-product” of (Polunchenko & Tartakovsky, 2010, Lemma 1) is that the quantity $\text{STADD}(\mathcal{S}_A^r)$ turns out to also provide a universal lowerbound on the unknown value of $\min_{T \in \Delta(\gamma)} \text{SADD}(T)$, and this lowerbound is valid for any $\gamma > 1$ and $r \geq 0$ such that $\mathcal{S}_A^r \in \Delta(\gamma)$; see (Polunchenko & Tartakovsky, 2010, Theorem 1). Specifically, introducing $\underline{\text{SADD}}(\mathcal{S}_A^r) \equiv \text{STADD}(\mathcal{S}_A^r)$, the following double inequality holds:

$$\underline{\text{SADD}}(\mathcal{S}_A^r) \leq \min_{T \in \Delta(\gamma)} \text{SADD}(T) \leq \text{SADD}(\mathcal{S}_A^r), \quad (9)$$

for any $A > 0$ and $r \geq 0$ such that $\text{ARL}(\mathcal{S}_A^r) \geq \gamma$, and any given $\gamma > 1$; cf. (Moustakides et al., 2011, Inequality (2.12), p. 579).

A few important comments are now in order:

1. On the one hand, the double inequality (9), namely, its left part, implies that the lowerbound $\underline{\text{SADD}}(\mathcal{S}_A^r) \equiv \text{STADD}(\mathcal{S}_A^r)$, where $\text{STADD}(T)$ is defined in (8), can be used as a benchmark to get an idea as to how much room there is for improvement in the way of SADD for a chart of interest. Should it so happen that the SADD of the chart of interest with the ARL to false alarm level set to $\gamma > 1$ is only a tiny bit greater than $\underline{\text{SADD}}(\mathcal{S}_A^r)$ assuming $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$, then the chart is almost minimax optimal in the sense of Pollak (1985).
2. On the other hand, the double inequality (9) also suggests the following optimization strategy for the GSR chart: for a given $\gamma > 1$, pick the chart's detection threshold $A > 0$ and headstart $r \geq 0$ in such a way so as to make the difference $\text{SADD}(\mathcal{S}_A^r) - \underline{\text{SADD}}(\mathcal{S}_A^r)$ as close to zero as is possible without violating the inequality $\text{ARL}(\mathcal{S}_A^r) \geq \gamma$. More formally, the optimal detection threshold A^* and headstart r^* values are to be selected as follows:

$$(r^*, A^*) = \arg \min_{r, A \geq 0} \{ \text{SADD}(\mathcal{S}_A^r) - \underline{\text{SADD}}(\mathcal{S}_A^r) \}, \text{ but } \text{ARL}(\mathcal{S}_A^r) = \gamma, \quad (10)$$

where $\gamma > 1$ is given; it goes without saying that both A^* and r^* are functions of $\gamma > 1$. The foregoing optimization strategy is originally due to Moustakides et al. (2011), and, in this work, we shall adapt it as well.

3. As we shall demonstrate in the next section, if the GSR chart's detection threshold A and initial score r are set to A^* and r^* , respectively, where A^* and r^* are as in (10) with $\gamma > 1$ given, then, conditional on $\text{ARL}(\mathcal{S}_A^r) = \gamma$, the difference $\text{SADD}(\mathcal{S}_A^r) - \underline{\text{SADD}}(\mathcal{S}_A^r)$ is nearly zero, even if $\gamma > 1$ is on the order of hundreds. Therefore, the GSR chart's third-order asymptotic optimality (5) does not necessarily require γ to be large.

The constrained optimization problem (10) can be solved numerically, e.g., with the aid of the numerical method proposed by Moustakides et al. (2011) and subsequently improved upon by Polunchenko et al. (2014b, 2014a). This is precisely the object of the next section.

3 Experimental Results

The plan now is to employ the numerical framework of Moustakides et al. (2011) and its improved version due to Polunchenko et al. (2014b, 2014a), and analyze the performance of the GSR chart given by (2) and (3) under different parameter

settings, including (and especially) the optimal choice given by the solution of the constrained optimization problem (10).

We begin with an examination of the level of the ARL to false alarm, i.e., $ARL(\mathcal{S}_A^r)$, treated as a function of the headstart $r \geq 0$, the detection threshold $A > 0$, and the magnitude of the change in the mean $\mu \neq 0$. With regard to the latter, for lack of space, let us consider only two cases: $\mu = 0.2$ and $\mu = 0.5$. The former may be considered a faint change, while the latter is a moderate change. Figures 1 depict $ARL(\mathcal{S}_A^r)$ as a function of $r \in [0, A]$ and $A \in [0, 1000]$. Specifically, Figure 1a is for $\mu = 0.2$ and Figure 1b is for $\mu = 0.5$. As can be seen from either figure, the bivariate function $ARL(\mathcal{S}_A^r)$ is almost linear in A (with r fixed) as well as in r (with A fixed). This is in perfect agreement with the aforementioned fact that $ARL(\mathcal{S}_A^r) \approx A/\xi - r$ where ξ is given by (4). Since, according to (Woodroffe, 1982, Table 3.1, p. 33), the value of ξ for $\mu = 0.2$ is roughly 0.89004 versus approximately 0.74762 for $\mu = 0.5$, the sensitivity of the ARL to false alarm level to the detection threshold is higher, the stronger the change. Figures 1 also include contours (shown as bold dark curves) corresponding the different fixed levels $\gamma > 1$ of the ARL to false alarm. Specifically, each of the contours is the solution set (r, A) of the equation $ARL(\mathcal{S}_A^r) = \gamma$ for the appropriate value of $\gamma = \{100, 200, \dots, 900, 1000\}$. These contours are important because the process of optimization of the GSR chart begins with picking a value for $\gamma > 1$, and then, with $\gamma > 1$ set and fixed, restricting attention to only those values of $A > 0$ and $r \geq 0$ for which the constraint $ARL(\mathcal{S}_A^r) = \gamma$ is satisfied. Due to space limitations, in this work we shall consider only three values of γ , namely, $\gamma = \{100, 500, 1000\}$.

Let us next look at Figures 2 and 3 which show $ADD_k(\mathcal{S}_A^r)$ as a function of $r \geq 0$ and $k = 0, 1, 2, \dots$ under the constraint $ARL(\mathcal{S}_A^r) = \gamma$ with $\gamma = \{100, 500, 1000\}$. Specifically, Figures 2 assume $\mu = 0.2$ while Figures 3 assume $\mu = 0.5$. With regard to the level $\gamma > 1$ of the ARL to false alarm, Figures 2a and 3a assume $\gamma = 100$, Figures 2b and 3b are for $\gamma = 500$, and Figures 2c and 3c assume $\gamma = 1000$. There are two important observations to make from either set of figures. First, it is evident that giving the SR chart a positive headstart equips the chart with the Fast Initial Response (FIR) feature: the chart becomes more sensitive to initial out-of-control situations. However, the flip side of the FIR feature is that the chart gets slower in situations when the process is initially in control but goes out of control later. It is worth reiterating that in order to retain the level of the ARL to false alarm assigning a higher value to the headstart is offset by an appropriate upward adjustment of the control limit. The second observation is that the maximal ADD, i.e., $SADD(\mathcal{S}_A^r) \triangleq \max_{0 \leq k < \infty} ADD_k(\mathcal{S}_A^r)$, is a sophisticated function of r , and the specific value of k at which the maximum is attained is hard to predict. As an aside, it is worth pointing out that the convergence of the ADDs to the steady-state regime is faster for $\mu = 0.5$ than for $\mu = 0.2$, which is consistent with one's intuition.

To better illustrate the FIR feature at work, let us look at Figures 4 and 5, which are effectively the projections of the 3D surfaces shown in Figures 2 and 3 onto the $(k, ADD_k(\mathcal{S}_A^r))$ -plane, made for a selection of values of r . Specifically, Figures 4 assume $\mu = 0.2$ and Figures 5 are for $\mu = 0.5$. The corresponding levels $\gamma > 1$ of the ARL to false alarm are given in the figures' subtitles. The figures clearly

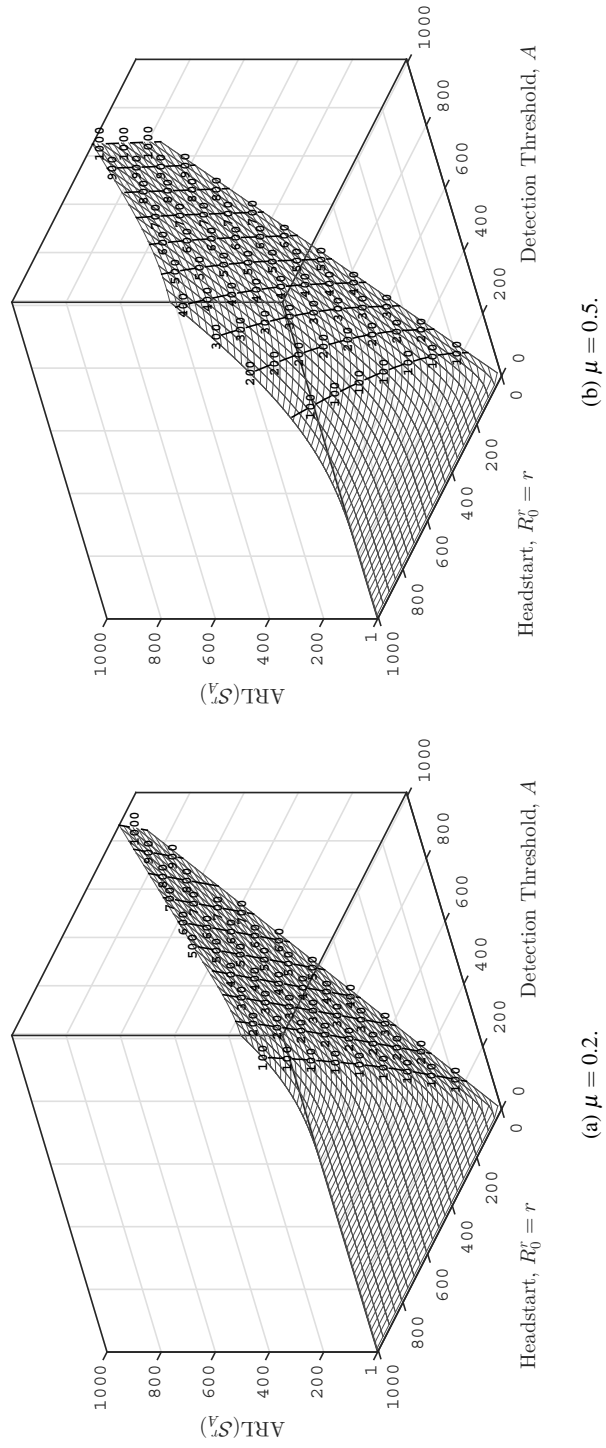


Fig. 1: $ARL(\mathcal{S}_A^r)$ as a function of the headstart $R_0^r = r \geq 0$ and the detection threshold $A > 0$ for $\mu = \{0.2, 0.5\}$.

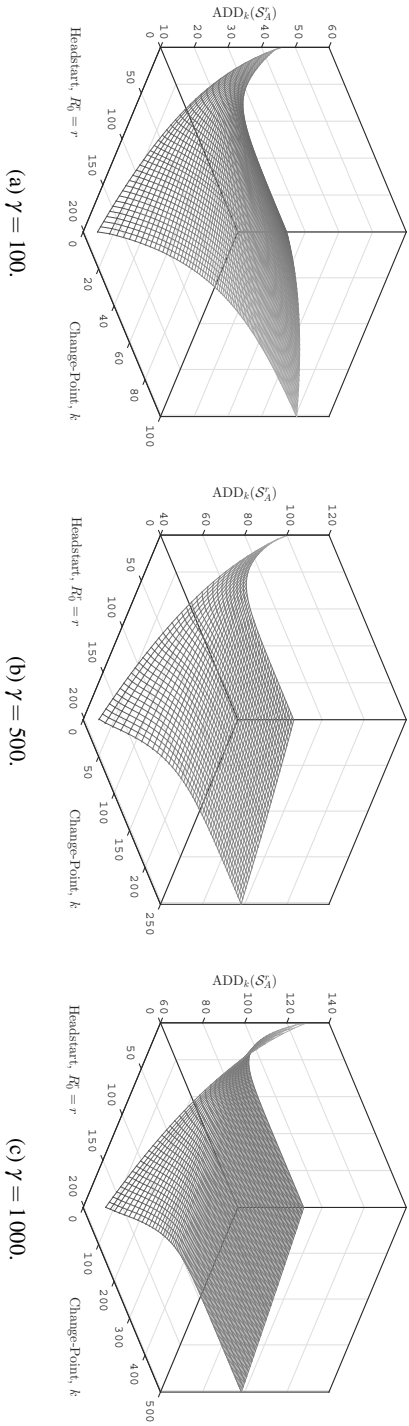


Fig. 2: $\text{ADD}_k(\mathcal{S}_A^r)$ as a function of the headstart $R_0^r = r \geq 0$, the change-point $k = 0, 1, \dots$, and the ARL to false alarm level $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$ for $\mu = 0.2$.

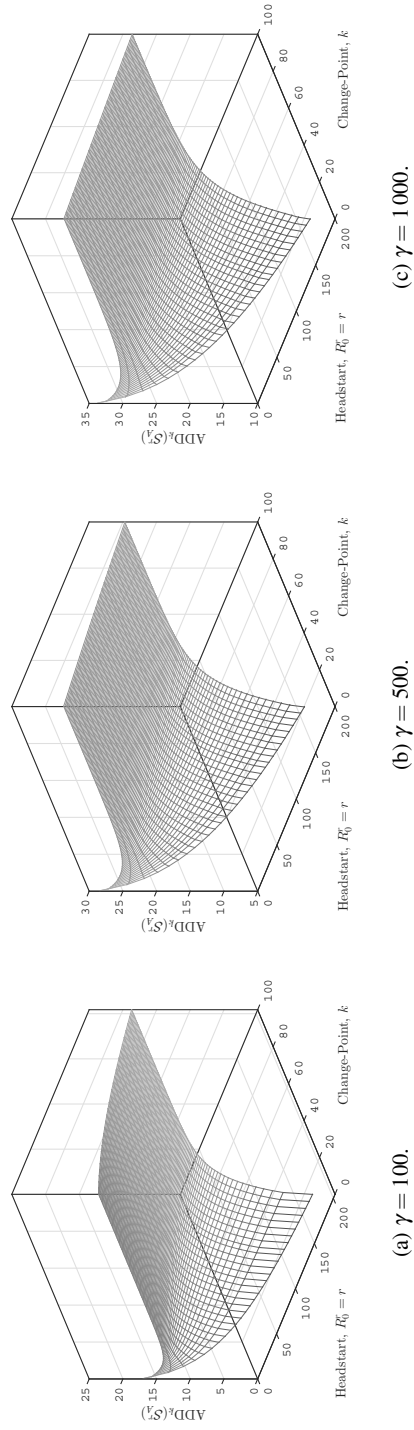


Fig. 3: $\text{ADD}_k(\mathcal{S}_A^r)$ as a function of the headstart $R_0^r = r \geq 0$, the change-point $k = 0, 1, \dots$, and the ARL to false alarm level $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$ for $\mu = 0.5$.

demonstrate that, as the headstart increases, the performance of the GSR chart for initial of early out-of-control situation improves. However, the performance in situations when the process goes out of control later degrades. The interesting question is whether it is possible to optimize this tradeoff. This question is hard to answer properly without getting the lowerbound $\text{SADD}(\mathcal{S}_A^r)$ involved, as is done in Figures 6 and 7.

Specifically, Figures 6 and 7 provide an idea as to the manner in which $\text{SADD}(\mathcal{S}_A^r)$ and $\underline{\text{SADD}}(\mathcal{S}_A^r)$ each depend on the headstart, assuming, as before, that every change in the headstart is accompanied by the appropriate adjustment of the detection threshold, so that the ARL to false alarm constraint is kept intact. More specifically, Figures 6 correspond to $\mu = 0.2$ and Figures 7 are for $\mu = 0.5$. The respective levels γ of the ARL to false alarm are again given in the subtitles.

It is evident from the figures that, regardless of the contrastness of the shift in the mean $\mu \neq 0$ and no matter the ARL to false alarm level $\gamma > 1$, the lowerbound is an upward arching smooth function of the initial score, and it has a distinct maximum. The figures also clearly indicate that the maximal ADD as a function of r has a minimum with the appearance of a down pointing cusp; the cusp is an indication that the way the maximal element of the sequence $\{\text{ADD}_k(\mathcal{S}_A^r)\}_{k \geq 0}$ and its location within the sequence depend on the headstart is highly nonlinear. The essential observation is that the lowerbound appears to peak at approximately the same (slightly smaller actually) headstart value as that at which the maximal ADD is minimized. Moreover, although the maximal ADD's minimum is higher than the lowerbound's maximum, the difference is not practically significant, even if γ is as low as 100, and is smaller, the higher the value of γ . Therefore, any other chart with the same level of the ARL to false alarm cannot possibly detect the shift in the mean with a detection delay substantially lower than that delivered by the optimized GSR chart, especially if the shift in the mean is contrast.

To draw a line under this section, in Tables 1 and 2, we give the optimal headstart and detection threshold values that have been computed by solving the constrained optimization problem (10) for $\gamma = \{100, 200, \dots, 900, 1000\}$ and $\mu = \{0.1, 0.2, \dots, 0.9, 1.0\}$. Recall also that our data model is symmetric with respect to the sign of $\mu \neq 0$. The tables also include the corresponding $\text{SADD}(\mathcal{S}_A^r)$ and $\underline{\text{SADD}}(\mathcal{S}_A^r)$ values. One can see from the tables that $\text{SADD}(\mathcal{S}_A^r) \approx \underline{\text{SADD}}(\mathcal{S}_A^r)$, which is to say that the detection capabilities of the optimized GSR chart are almost the best. One can also see that the effect of headstarting is the stronger, the fainter the anticipated shift in the mean. If the latter is fairly contrast, the optimal headstart value is close to zero. In addition, the tables suggest that the optimal headstart value, as a function of the ARL to false alarm level $\gamma > 1$, has a *finite* limit as $\gamma \rightarrow +\infty$; the convergence to the limiting value is the slower, the weaker the change. However, a closed-form formula for this limiting value is prohibitively difficult to obtain.

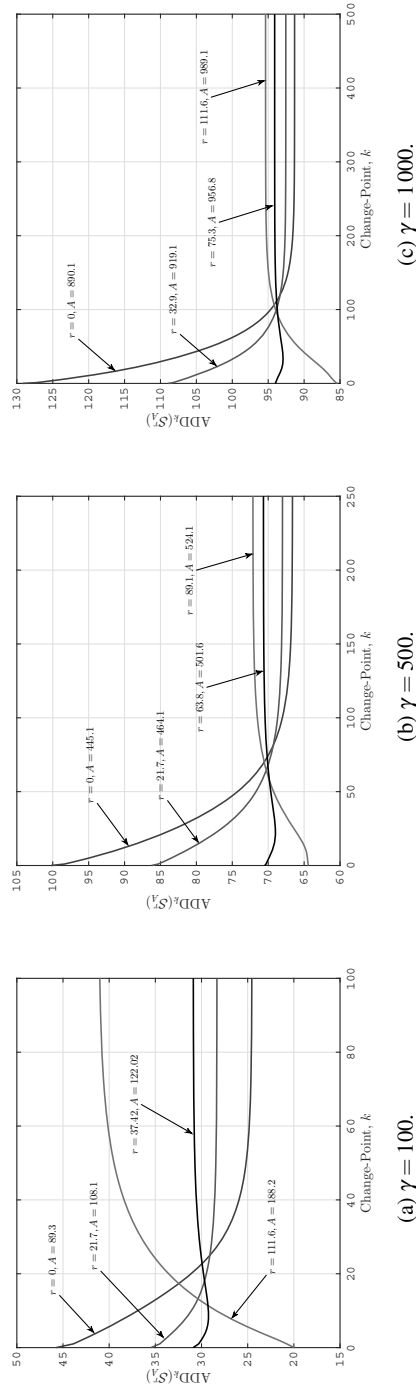


Fig. 4: $ADD_k(\mathcal{S}_A^r)$ as a function of the headstart $R_0^r = r \geq 0$, the change-point $k = 0, 1, \dots$, and the ARL to false alarm level $ARL(\mathcal{S}_A^r) = \gamma > 1$ for $\mu = 0.2$.

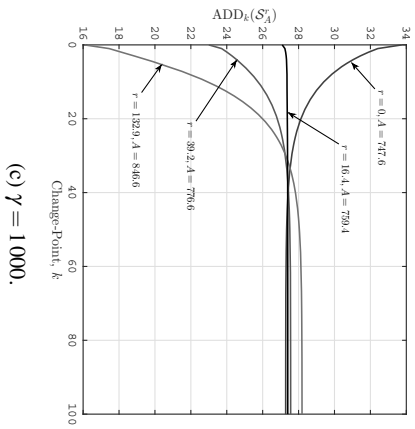
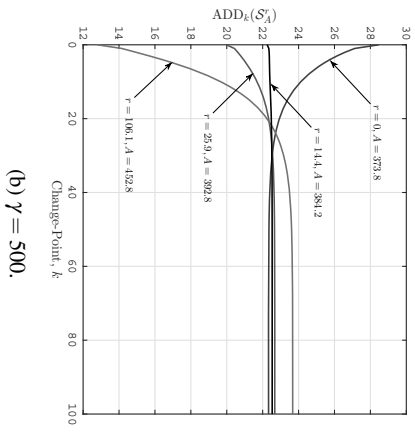
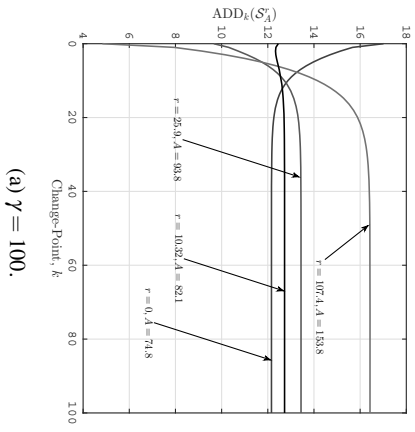


Fig. 5: $\text{ADD}_k(\mathcal{S}_A^{(r)})$ as a function of the headstart $R'_0 = r \geq 0$, the change-point $k = 0, 1, \dots$, and the ARL to false alarm level $\text{ARL}(\mathcal{S}_A^{(r)}) = \gamma > 1$ for $\mu = 0.5$.

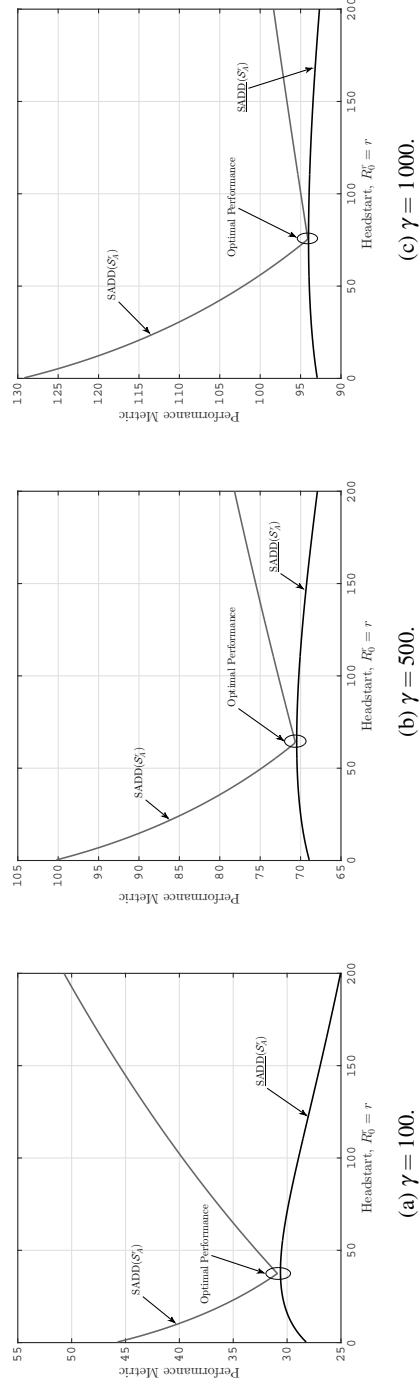
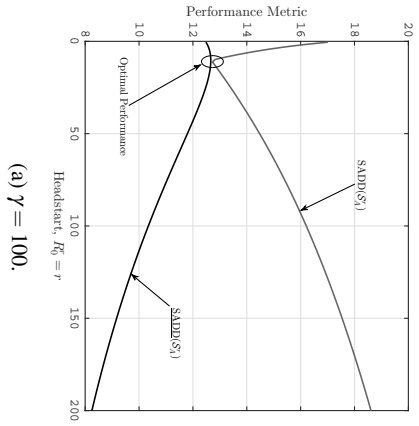
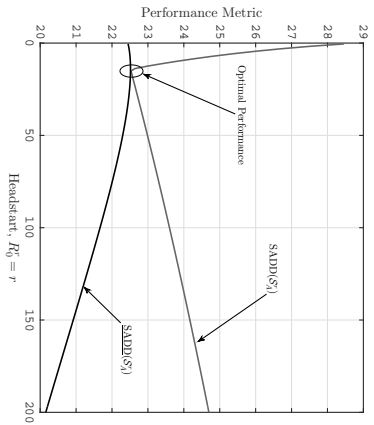


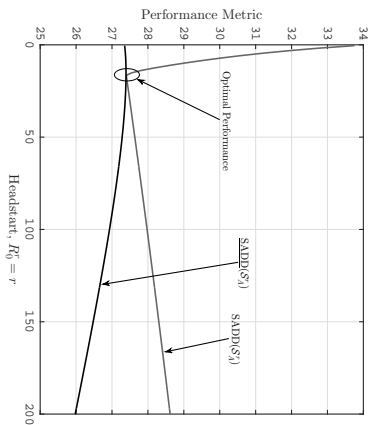
Fig. 6: $\text{SADD}(\mathcal{S}_A^r)$ and $\text{SADD}(\mathcal{S}_A^r)$ as functions of the headstart $R_0^r = r$ and the ARL to false alarm level $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$ for $\mu = 0.2$.



(a) $\gamma = 100$.



(b) $\gamma = 500$.



(c) $\gamma = 1000$.

Fig. 7: $\text{SADD}(\mathcal{S}_A^r)$ and $\text{SADD}(\mathcal{S}_A^r)$ as functions of the headstart $R_0^r = r$ and the ARL to false alarm level $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$ for $\mu = 0.5$.

Table 1: Optimal headstart, $r^* > 0$, control limit, $A^* > 0$, maximal ADD, $\text{SADD}(\mathcal{S}_A^r)$, and the lowerbound, $\text{SADD}(\mathcal{S}_A^r)$, as functions of the shift magnitude, $\mu > 0$, and the ARL to false alarm level, $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$, for $\gamma = \{100, 200, 300, 400, 500\}$.

$\text{ARL}(\mathcal{S}_A^r) = \gamma$	Performance Characteristic	Change Magnitude ($\mu > 0$)									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
100	r^*	83.93	37.42	21.96	14.53	10.32	7.66	5.89	4.64	3.74	3.05
	A^*	173.25	122.02	102.11	90.43	82.14	75.6	70.14	65.38	61.14	57.31
	$\text{SADD}(\mathcal{S}_A^r)$	49.65	30.9	21.6	16.17	12.68	10.28	8.55	7.25	6.25	5.46
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	48.76	30.62	21.49	16.13	12.66	10.27	8.54	7.25	6.25	5.46
200	r^*	114.43	48.29	27.23	17.49	12.14	8.86	6.72	5.27	4.24	3.48
	A^*	296.37	220.71	190.51	172	158.26	147	137.28	128.63	120.79	113.58
	$\text{SADD}(\mathcal{S}_A^r)$	79.79	45.39	30.1	21.75	16.61	13.19	10.79	9.03	7.69	6.65
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	78.7	45.14	30.03	21.72	16.6	13.19	10.79	9.03	7.69	6.65
300	r^*	135.53	55.1	30.29	19.12	13.1	9.54	7.27	5.7	4.6	3.77
	A^*	410.61	315.77	277.06	252.52	233.74	218.04	204.24	191.76	180.34	169.78
	$\text{SADD}(\mathcal{S}_A^r)$	103.23	55.71	35.87	25.41	19.13	15.04	12.19	10.13	8.58	7.38
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	102.08	55.5	35.81	25.4	19.13	15.03	12.19	10.13	8.58	7.38
400	r^*	151.87	60.02	32.41	20.2	13.81	10.05	7.65	6.01	4.83	3.98
	A^*	520.37	409.15	362.81	332.61	309.04	288.95	271.07	254.81	239.82	225.94
	$\text{SADD}(\mathcal{S}_A^r)$	122.8	63.86	40.29	28.17	21.02	16.4	13.22	10.93	9.22	7.91
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	121.65	63.68	40.25	28.16	21.01	16.39	13.22	10.93	9.22	7.91
500	r^*	165.27	63.84	33.98	21.03	14.36	10.45	7.95	6.25	5.03	4.14
	A^*	627.35	501.56	448.1	412.5	384.21	359.78	337.86	317.81	299.29	282.07
	$\text{SADD}(\mathcal{S}_A^r)$	139.75	70.63	43.86	30.39	22.52	17.48	14.05	11.57	9.73	8.33
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	138.63	70.48	43.86	30.39	22.52	17.48	14.03	11.57	9.73	8.33

Table 2: Optimal headstart, $r^* > 0$, control limit, $A^* > 0$, maximal ADD, $\text{SADD}(\mathcal{S}_A^r)$, and the lowerbound, $\text{SADD}(\mathcal{S}_A^r)$, as functions of the shift magnitude, $\mu > 0$, and the ARL to false alarm level, $\text{ARL}(\mathcal{S}_A^r) = \gamma > 1$, for $\gamma = \{600, 700, 800, 900, 1000\}$.

$\text{ARL}(\mathcal{S}_A^r) = \gamma$	Performance Characteristic	Change Magnitude ($\mu > 0$)									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
600	r^*	176.63	66.94	35.24	21.73	14.81	10.78	8.19	6.45	5.2	4.28
	A^*	732.41	593.32	533.12	492.28	459.31	430.56	404.61	380.8	358.73	338.18
	$\text{SADD}(\mathcal{S}_A^r)$	153.8	76.46	46.95	32.26	23.77	18.38	14.71	12.09	10.15	8.67
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	153.71	76.33	46.92	32.25	23.77	18.37	14.71	12.09	10.15	8.67
700	r^*	186.49	69.53	36.25	22.32	15.21	11.06	8.42	6.61	5.34	4.39
	A^*	836.06	684.63	617.94	571.98	534.37	501.32	471.35	443.76	418.15	394.28
	$\text{SADD}(\mathcal{S}_A^r)$	168.37	81.58	49.59	33.87	24.85	19.15	15.29	12.54	10.51	8.96
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	167.32	81.47	49.57	33.86	24.85	19.14	15.28	12.54	10.51	8.96
800	r^*	195.2	71.73	37.14	22.84	15.55	11.3	8.6	6.77	5.47	4.5
	A^*	938.62	775.59	702.67	651.62	609.38	572.04	538.06	506.71	477.58	450.38
	$\text{SADD}(\mathcal{S}_A^r)$	180.76	86.16	51.94	35.29	25.79	19.82	15.79	12.93	10.82	9.22
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	179.75	86.06	51.92	35.28	25.79	19.82	15.79	12.93	10.82	9.22
900	r^*	202.99	73.65	37.93	23.31	15.86	11.53	8.78	6.91	5.57	4.59
	A^*	1040.31	866.3	787.3	731.22	684.37	642.75	604.77	569.66	536.98	506.46
	$\text{SADD}(\mathcal{S}_A^r)$	192.18	90.3	54.04	36.55	26.64	20.42	16.24	13.28	11.1	9.44
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	191.21	90.21	54.03	36.55	26.63	20.42	16.24	13.28	11.1	9.44
1 000	r^*	210.04	75.34	38.62	23.71	16.14	11.73	8.93	7.04	5.68	4.66
	A^*	1141.3	956.81	871.86	810.77	759.35	713.44	671.46	632.6	596.38	562.54
	$\text{SADD}(\mathcal{S}_A^r)$	202.79	94.09	55.96	37.7	27.39	20.96	16.64	13.59	11.35	9.65
	$\underline{\text{SADD}}(\mathcal{S}_A^r)$	201.86	94.01	55.94	37.69	27.39	20.95	16.64	13.59	11.35	9.64

4 Concluding Remarks

In summary we see that

1. Starting an SR chart off a nonzero initial score lessens the ARL to false alarm, so that the chart's in-control performance is worse than when no headstart is used. On the flip side, however, the chart becomes more sensitive to initial out-of-control situations. This is precisely the FIR phenomenon.
2. The drop in the ARL to false alarm caused by a positive headstart value can be compensated by an increase of the control limit. While this would negatively affect the chart's out-of-control performance, the magnitude of the effect appears to be not substantial.
3. The FIR feature comes at the price of poorer performance in situations when the process under surveillance is initially in control but goes out of control later. In particular, if the process is not expected to shift out of control for a long while, then no headstarting is necessary, because the SR chart's steady-state performance would degrade otherwise.

The same observations were previously made by Lucas & Crosier (1982) about the CUSUM chart.

Our additional and more important contribution consists in a deeper investigation of the headstart-vs-control-limit tradeoff: the overall performance of the GSR chart optimized not only with respect to the headstart but also with respect to the control limit proved to be nearly the best one can get amid complete uncertainty as to when the observed process may go out of control. This is a direct implication of the GSR chart's strong optimality properties established by Pollak & Tartakovsky (2009), Shiryaev & Zryumov (2009), Tartakovsky & Polunchenko (2010), Polunchenko & Tartakovsky (2010), and by Tartakovsky et al. (2012). The optimal headstart and control limit values, and the corresponding out-of-control performance and its lowerbound, for a variety of cases, are given in Tables 1 and 2.

The benefits of optimizing the GSR chart are the greater, the fainter the change. From a practical standpoint, this means that if one is interested in detecting a faint change, then the GSR chart with optimally selected control limit and headstart is the way to go. The size of the actual efficiency improvement can be estimated using Tables 1 and 2. However, if the anticipated change to be detected is more or less contrast, then the GSR chart, whether optimized or not, will not offer any substantial advantage (in terms of the speed of detection) over the CUSUM scheme or the EWMA chart.

Acknowledgements The author's effort was partially supported by the Simons Foundation via a Collaboration Grant in Mathematics under Award # 304574.

References

- Girshick MA and Rubin H (1952). A Bayes approach to a quality control model. *Ann Math Statist* 23(1), 114-125. doi:[10.1214/aoms/1177729489](https://doi.org/10.1214/aoms/1177729489).
- Kenett R and Pollak M (1986). A semi-parametric approach to testing for reliability growth, with application to software systems. *IEEE Trans Rel* 35(3), 304-311. doi:[10.1109/TR.1986.4335439](https://doi.org/10.1109/TR.1986.4335439).
- Kenett R and Pollak M (1996). Data-analytic aspects of the Shiryaev-Roberts control chart: Surveillance of a non-homogeneous Poisson process. *J Appl Stat* 23(1), 125-138. doi:[10.1080/02664769624413](https://doi.org/10.1080/02664769624413).
- Lucas JM and Crosier RB (1982). Fast initial response for CUSUM quality-control schemes: Give your CUSUM a head start. *Technometrics* 24(3), 199-205. doi:[10.2307/1268679](https://doi.org/10.2307/1268679).
- Mevorach Y and Pollak M (1991). A small sample size comparison of the CUSUM and Shiryaev-Roberts approaches to changepoint detection. *Amer J Math Management Sci* 11(3&4), 277-298. doi:[10.1080/01966324.1991.10737312](https://doi.org/10.1080/01966324.1991.10737312).
- Moustakides GV, Polunchenko AS and Tartakovsky AG (2011). A numerical approach to performance analysis of quickest change-point detection procedures. *Statist Sinica* 21(2), 571-596.
- Moustakides GV, Polunchenko AS and Tartakovsky AG (2009). Numerical comparison of CUSUM and Shiryaev-Roberts procedures for detecting changes in distributions. *Commun Stat Theory Methods* 38(16), 3225-3239.
- Page ES (1954). Continuous inspection schemes. *Biometrika* 41(1&2), 100-115. doi:[10.2307/2333009](https://doi.org/10.2307/2333009).
- Pepelyshev A and Polunchenko AS (2016). Real-time financial surveillance via quickest change-point detection methods. *Stat Interface* (in press). Available via ArXiv: <https://arxiv.org/abs/1509.01570>. Cited 3 Jul 2016.
- Pollak M (1985). Optimal detection of a change in distribution. *Ann Statist* 13(1), 206-222. doi:[10.1214/aos/1176346587](https://doi.org/10.1214/aos/1176346587).
- Pollak M (1987). Average Run Lengths of an optimal method of detecting a change in distribution. *Ann Statist* 15(2), 749-779. doi:[10.1214/aos/1176350373](https://doi.org/10.1214/aos/1176350373).
- Pollak M (2009). The Shiryaev-Roberts changepoint detection procedure in retrospect – Theory and practice. In: *Proce 2nd Int'l Workshop Sequential Methodol*, University of Technology of Troyes, Troyes, France, 15-17 Jun, 2009.
- Pollak M and Tartakovsky AG (2009). Optimality properties of the Shiryaev-Roberts procedure. *Statist Sinica* 19, 1729-1739.
- Polunchenko AS and Sokolov G (2014). Toward optimal design of the Generalized Shiryaev-Roberts procedure for quickest change-point detection under exponential observations. In: *Proc 2014 Int'l Conf Engineering & Telecommunications*, Moscow Institute of Physics and Technology, Moscow, Russia, 26-28 Nov 2014; pp. 51-55. doi:[10.1109/EnT.2014.37](https://doi.org/10.1109/EnT.2014.37).
- Polunchenko AS, Sokolov G and Du W (2014). Efficient performance evaluation of the generalized Shiryaev-Roberts detection procedure in a multi-cyclic setup. *Appl Stoch Models Bus Ind*, 30(6), 723-739. doi:[10.1002/asmb.2026](https://doi.org/10.1002/asmb.2026).

- Polunchenko AS, Sokolov G and Du W (2014). An accurate method for determining the pre-change Run-Length distribution of the Generalized Shiryaev-Roberts detection procedure. *Sequential Anal* 33(1), 112-134. doi:[10.1080/07474946.2014.856642](https://doi.org/10.1080/07474946.2014.856642).
- Polunchenko AS, Sokolov G and Du W (2016). On robustness of the Shiryaev-Roberts change-point detection procedure under parameter misspecification in the post-change distribution. *Commun Stat Simul Comput*, (in press). doi:[10.1080/03610918.2015.1039131](https://doi.org/10.1080/03610918.2015.1039131). Available via ArXiv: <https://arxiv.org/abs/1504.04722>. Cited 3 Jul 2016.
- Polunchenko AS and Tartakovsky AG (2010). On optimality of the Shiryaev-Roberts procedure for detecting a change in distribution. *Ann Statist* 38(6), 3445-3457. doi:[10.1214/09-AOS775](https://doi.org/10.1214/09-AOS775).
- Polunchenko AS and Tartakovsky AG (2012). State-of-the-art in sequential change-point detection. *Methodol Comput Appl Probab* 44(3), 649-684. doi:[10.1007/s11009-011-9256-5](https://doi.org/10.1007/s11009-011-9256-5).
- Roberts SW (1959). Control chart tests based on geometric moving averages. *Technometrics* 1(3), 239-250. doi:[10.2307/1271439](https://doi.org/10.2307/1271439).
- Roberts SW (1966). A comparison of some control chart procedures. *Technometrics* 8(3), 411-430. doi:[10.2307/1266688](https://doi.org/10.2307/1266688).
- Shiryaev AN (1961). The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Math Dokl* 2, 795-799.
- Shiryaev AN (1963). On optimum methods in quickest detection problems. *Theory of Probab Appl* 8(1), 22-46. doi:[10.1137/1108002](https://doi.org/10.1137/1108002).
- Shiryaev AN and Zryumov PY (2009). On the linear and nonlinear generalized Bayesian disorder problem (discrete time case). In: *Optimality and Risk – Modern Trends in Mathematical Finance. The Kabanov Festschrift*, Delbaen F, Rásonyi M and Stricker Ch (eds), pp. 227-235, Springer-Verlag, Berlin, Germany. doi:[10.1007/978-3-642-02608-9_12](https://doi.org/10.1007/978-3-642-02608-9_12).
- Tartakovsky AG and Ivanova IV (1992). Comparison of some sequential rules for detecting changes in distributions. *Probl Inf Transm* 28(2), 117-124.
- Tartakovsky AG, Pollak M and Polunchenko AS (2012). Third-order asymptotic optimality of the Generalized Shiryaev-Roberts changepoint detection procedure. *Theory Probab Appl* 56(3), 457-484. doi:[10.1137/S0040585X97985534](https://doi.org/10.1137/S0040585X97985534).
- Tartakovsky AG, Polunchenko AS and Sokolov G (2013). Efficient computer network anomaly detection by changepoint detection methods. *IEEE J Sel Top Sign Proces* 7(1), 4-11. doi:[10.1109/JSTSP.2012.2233713](https://doi.org/10.1109/JSTSP.2012.2233713).
- Tartakovsky AG and Polunchenko AS (2010). Minimax optimality the Shiryaev-Roberts procedure. In: *Proc 5th Int'l Workshop Applied Probab*, Universidad Carlos III de Madrid, Colmenarejo Campus, Spain, 5-8 Jul, 2010.
- Tartakovsky AG, Polunchenko AS and Moustakides GV (2009). Design and comparison of Shiryaev-Roberts- and CUSUM-type change-point detection procedures. In: *Proc 2nd Int'l Workshop Sequential Methodol*, University of Technology of Troyes, Troyes, France, 15-17 Jun, 2009.
- Woodroffe M (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia, PA.